Understanding asthma phenotypes: the World Asthma Phenotypes (WASP) international collaboration

Lucy Pembrey[1], Mauricio L. Barreto[2,3], Jeroen Douwes[4], Philip Cooper[5,6,7], John Henderson[8], Harriet Mpairwe[9], Cristina Ardura-Garcia[7], Martha Chico[7], Collin Brooks[4], Alvaro A. Cruz[10,11], Alison M. Elliott[9], Camila A. Figueiredo[2], Sinéad M. Langan[1], Beatrice Nassanga[9], Susan Ring[8], Laura Rodrigues[1] and Neil Pearce[1,4]

Affiliations: [1]London School of Hygiene and Tropical Medicine, London, UK. [2]Institute of Collective Health, Federal University of Bahia, Salvador, Brazil. [3]Center for Data and Knowledge Integration for Health (CIDACS), Fiocruz, Salvador, Brazil. [4]Centre for Public Health Research, Massey University, Wellington, New Zealand. [5]St George's University of London, London, UK. [6]Universidad Internacional del Ecuador, Quito, Ecuador. [7]Fundacion Ecuatoriana Para Investigacion en Salud, Quito, Ecuador. [8]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [9]Medical Research Council (MRC)/Uganda Virus Research Institute (UVRI) Uganda Research Unit on AIDS, Entebbe, Uganda. [10]The Programme for Control of Asthma and Allergic Rhinitis (ProAR), Federal University of Bahia, Salvador, Brazil. [11]Institute for Health Sciences, Federal University of Bahia, Salvador, Brazil.

Correspondence: Neil Pearce, Dept of Medical Statistics and Centre for Global Chronic Conditions, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: neil.pearce@lshtm.ac.uk

ABSTRACT   The World Asthma Phenotypes (WASP) study started in 2016 and has been conducted in five centres, in the UK, New Zealand, Brazil, Ecuador and Uganda.

The objectives of this study are to combine detailed biomarker and clinical information in order to 1) better understand and characterise asthma phenotypes in high-income countries (HICs) and low and middle-income countries (LMICs), and in high and low prevalence centres; 2) compare phenotype characteristics, including clinical severity; 3) assess the risk factors for each phenotype; and 4) assess how the distribution of phenotypes differs between high prevalence and low prevalence centres.

Here we present the rationale and protocol for the WASP study to enable other centres around the world to carry out similar analyses using a standardised protocol. Large collaborative and integrative studies like this are essential to further our understanding of asthma phenotypes. The findings of this study will help elucidate the aetiological mechanisms of asthma and might potentially identify new causes and guide the development of new treatments, thereby enabling better management and prevention of asthma in both HICs and LMICs.

@ERSpublications
A protocol for the World Asthma Phenotypes study to better understand and characterise asthma phenotypes in five centres in the UK, New Zealand, Brazil, Ecuador and Uganda http://ow.ly/71CR30ldsU4

Introduction

It is now well established that there are multiple phenotypes and endotypes of asthma [1, 2]. 15 to 20 years ago it was widely believed that asthma was an allergic/atopic disease caused by allergen exposure. More recent work has shown that this model does not explain the global asthma patterns and time trends [3–8].

A recent Lancet Commission report on asthma called for greater recognition of the various phenotypes, with different underlying pathological mechanisms, often grouped under the non-specific label of asthma. The authors acknowledge the need for better research to identify different traits, such that targeted treatments can be tailored to individual characteristics [2]. Here we define asthma phenotypes as the combination of clinical, demographic and pathological characteristics of asthma and we include pathological and inflammatory endotypes as a subset within this definition.

Characterisation of asthma phenotypes is important in order to: 1) better understand the aetiological mechanisms of asthma; 2) identify specific causes; 3) guide the development of new therapeutic measures that will be effective for all asthmatic patients; and 4) enable better management and prevention of asthma in both high-income countries (HICs) and low and middle-income countries (LMICs). There are major research opportunities to address these issues [9] due to the recent development and establishment of techniques for sample collection from the target organ, *i.e.* the airways (sputum induction and nasal lavage), novel analytical approaches for biological samples (microbiome assessment and epigenetics) and emerging biostatistical methods for integrating data from multiple sources and levels (bioinformatics). There is therefore an unprecedented potential to go beyond the old atopic/non-atopic [10], eosinophilia/ non-eosinophilia categorisation of asthma phenotypes.

For these reasons, the World Asthma Phenotypes (WASP) international collaboration has been formed to investigate and characterise asthma phenotypes in more depth. The study will initially be conducted in five countries, but we are publishing the background, justification and protocol here to enable other centres around the world to carry out similar analyses using a standardised protocol.

What is known about asthma phenotypes?

Until relatively recently, asthma was widely regarded, particularly in children, as an atopic disease involving allergen exposure, IgE-mediated sensitisation with a Type 2 T-helper cell (Th2) lymphocyte response and subsequent interleukin-5 (IL-5) mediated eosinophilic airways inflammation, resulting in reversible airflow obstruction and clinical symptoms (asthma). This assumption is increasingly being challenged and disproved [3, 11, 12], and there is growing interest in alternative inflammatory and non-inflammatory mechanisms for asthma [3, 13]. Studies using induced sputum have demonstrated that less than 50% of asthma cases are attributable to eosinophilic airway inflammation [3, 14]. Patients may have severe and persistent asthma in the absence of detectable eosinophilic inflammation and may experience exacerbations without an increase in eosinophilic inflammation [3]. Repeated assessments of airway inflammation have shown that the non-eosinophilic asthma (NEA) phenotype may be reproducible both in the short-term (4 weeks) and the long-term (1–5 years), but the evidence is equivocal [15, 16] and the underlying mechanisms are not fully understood.

The approaches to phenotyping asthma most commonly used to date include clinical, trigger-related, demographic and pathological factors. Most early attempts to define phenotypes were relatively one-dimensional [17] and based on simple classifications such as age of onset or atopic/non-atopic. More recently, a multi-dimensional approach has been increasingly used [17], involving statistical methods such as latent class analysis [18, 19]. However, while some recent studies have included many variables to identify asthma sub-types [20], most have continued to be based solely on clinical/demographic characteristics or pathological characteristics and few population-based studies have used multiple aspects to assemble comprehensive phenotypes. In the large Mechanisms of the Development of ALLergy (MeDALL) study, allergic multimorbidity phenotypes have been identified based on eczema, rhinitis and asthma; however, further characterisation of specific asthma phenotypes based on pathophysiological characteristics has not been performed [21].

Several features of the existing studies on asthma phenotypes are notable. First, and most importantly, nearly all of these studies have been conducted in HICs, with the exception of a small study in Brazil [22] which did not directly address the issue of phenotypes and a recent study from China [23]. The key strength of the study protocol presented here is that it is the first to undertake complex phenotype analyses in LMICs and to compare the results with those in HICs using the same methodology.

Secondly, there has been little progress on clinical wheezing phenotypes since the seminal paper of MARTINEZ *et al.* [24] in 1995, which identified the distinctions between transient wheeze, late-onset wheeze and persistent wheeze in the first 6 years of life. Most (but not all) of the clinical classifications of phenotypes are similar, in some instances with a few more added categories [25–30] (table S1). This is

perhaps not surprising, since most attempts to reclassify asthma phenotypes have only collected information on symptoms and atopy status and thus the resulting clinical phenotypes have been restricted to the categories that are possible using this limited information (atopic/non-atopic and early/late/persistent wheezing).

The need to characterise asthma phenotypes on a pathological rather than a clinical basis has long been recognised [31] and driven by the need to improve treatment, since some asthmatic patients respond to conventional treatment while others do not [2]. However, most studies have not made full use of techniques for sample collection from the target organ (sputum induction and nasal lavage), more novel methods of biomarker analysis (microbiome assessment, epigenetics and other "omics" technologies), or developing methods of biostatistical analysis (bioinformatics). Despite this, several studies have identified asthma phenotypes based on markers of airways inflammation (table S1). Perhaps the prototypical study attempting to characterise phenotypes based on such markers is that of SIMPSON et al. [32], which concludes that there are four inflammatory subtypes: eosinophilic asthma (EA), neutrophilic asthma, mixed granulocytic asthma and paucigranulocytic asthma. HALDAR et al. [33] have summarised and characterised these subtypes.

Other work has also characterised airways inflammation [34] but few studies have combined this pathological data with clinical information to produce a more comprehensive approach to describing asthma phenotypes. Until recently, the methods for sampling the airways have been invasive and therefore restricted to adult populations or opportunistic sampling of children. For this reason, there is little detailed population-based data regarding such phenotypes in children and adolescents, and most studies have been restricted to examining EA and NEA. It is also notable that few studies have found neutrophilic asthma to be important in children [35]; however, this issue has not been explored in LMICs, where non-atopic asthma appears to be relatively more important. The study protocol presented here is the first to address this issue in LMICs.

Neutrophilic asthma (often identified as <2% eosinophils and ⩾61% neutrophils in sputum [36]) appears to involve an IL-8 mediated neutrophil influx, with the subsequent neutrophil activation being a potent stimulus to increased airway hyper-responsiveness [6]. Although the relevant stimuli appear diverse (e.g. endotoxin, ozone, particulates and virus infection), the features are consistent with activation of innate immune mechanisms (involving Toll-like receptors and CD14) rather than (or in addition to) IgE-mediated activation of acquired immunity. There is also the potential for combined activation of both innate and allergen specific inflammatory mechanisms. This may be the case in mixed granulocytic asthma and may explain the ability of ozone and nitrogen dioxide (NO_2) to potentiate allergen-induced asthmatic responses, although the pathophysiological mechanisms are not clear.

Little is known about other phenotypes and, in particular, it is possible that in some cases NEA may involve neurogenic mechanisms which may interact with airways inflammation and remodelling [37]; however, these have rarely been studied, despite asthma being viewed predominantly as a neurological disorder until the second half of the 20th century [38].

These differences in phenotypes are important not only for understanding the aetiology of asthma, but also for asthma management. Although EA and NEA cases are generally prescribed the same asthma medication, there may be important differences in their responses to medications and other aspects of management [33]. In particular, subjects with NEA are reported to have a poor short-term response to treatment with inhaled corticosteroids (ICS), which are primarily intended to suppress eosinophilic inflammation. Furthermore, the two phenotypes have not been systematically compared with respect to their expressions of other potentially relevant biological markers (e.g. markers of neural involvement [39, 40] or the airways microbiome [41]).

A potential confounding factor when comparing asthma phenotypes between centres may be differences in patterns of ICS use between HICs and LMICs. In particular, in some LMICs it is unlikely that many participants will be using ICS, whilst a relatively high proportion of participants in HICs are expected to regularly be using ICS. This is important, as a number of studies in adults show that initiating ICS treatment or increasing ICS dose is associated with a significant reduction in both sputum eosinophil percentage and absolute number, and that a corresponding increase in eosinophils is observed when discontinuing ICS [15, 36, 42–45]. In some of these studies, a reduction in ICS dose has also been associated with an increase in the prevalence of the EA phenotype [15, 42]. However, change in inflammatory phenotype due to ICS treatment has not been a consistent finding and may be dependent on the population studied. For example, in a double-blind crossover study of asthmatic subjects with mild to moderate asthma, there was no significant difference in EA prevalence when comparing the same group after placebo or budesonide treatment [15]. There are also very limited data available in children and, while variation in inflammatory phenotypes over time was demonstrated in one study of 51 children with

severe asthma and 28 with mild-to-moderate asthma, there was no evidence that this was related to a change in either asthma control or ICS dose [16]. To our knowledge, no other studies have examined the effects of altering ICS treatment on inflammatory phenotypes in paediatric populations.

In summary, it is currently unclear whether asthma involves a collection of different conditions, or is a single condition with multiple mechanisms and phenotypes, or what the underlying phenotypes may be [2]. Much of the previous work was based in high-income/high-prevalence countries and has, in general, involved relatively limited information on biological markers beyond blood and sputum differential cell counts and serum proteins. Techniques for sample collection from the airways are increasingly available, as are alternative methods of analysis and new biostatistical methods for integrating data from multiple sources and levels. Thus, it is an ideal time to re-examine the identification of novel asthma phenotypes using detailed information from patients in a wide variety of geographical settings. In particular, a major shortcoming of work on asthma phenotypes to date is the scarcity of studies looking at markers of inflammation in sputum samples in LMICs and the lack of comparisons between LMICs and HICs.

The WASP international collaboration

The overarching objectives of the WASP study are as follows: 1) to better understand and characterise asthma phenotypes in HICs and LMICs, and in high and low prevalence centres; 2) to compare phenotype characteristics, including clinical severity; 3) to assess the risk factors for each phenotype; and 4) to assess how the distribution of phenotypes differs between high prevalence and low prevalence centres.

The key features of this study are the inclusion of: 1) both high and low prevalence centres from both HICs and LMICs; 2) more detailed and varied biomarker information than has been used for previous studies identifying asthma phenotypes; and 3) new bioinformatics methods for integrating data from multiple sources and levels.

Study protocol

We are conducting a multi-country study to examine asthma phenotypes in a variety of settings, including those with both high and low asthma prevalence, and from both HICs and LMICs. By identifying biological mechanisms that may or may not be common to asthma phenotypes in diverse contexts, the study aims to identify critical causal pathways relevant to both prevention and treatment.

Study centres

The study will be conducted in five centres, Bristol in the UK (the Avon Longitudinal Study of Parents and Children (ALSPAC)), Wellington in New Zealand, Salvador in Brazil, Quininde in Ecuador and

TABLE 1 Collaborating studies and centres		
Centre	**Characteristics**	**Study type**
Avon Longitudinal Study of Parents and Children (ALSPAC), Bristol, UK	A high prevalence centre in a HIC	Birth cohort study with extensive detailed longitudinal information[#] New data collection at age 26–27 years
Centre for Public Health Research (CPHR), Wellington, New Zealand	A high prevalence centre in a HIC	NZAZCS birth cohort New data collection at age 16–20 years Cross-sectional study in children aged 12–16 years recruited from schools and the community
Social Change, Asthma and Allergy in Latin America (SCAALA), Salvador, Brazil	A high prevalence centre in a LMIC	SCAALA cohort New data collection at age 11–19 years Cross-sectional study in children aged 11–19 years recruited from three schools
Social Change, Asthma and Allergy in Latin America (SCAALA), Esmereldas, Ecuador	A medium prevalence centre in a LMIC	Population-based cohort New data collection at age 8–12 years Cross-sectional study in children aged 12–16 years recruited from schools
Entebbe childhood asthma study, Entebbe, Uganda	A low prevalence centre in a LMIC	Cross-sectional study in children aged 12–16 years recruited from schools

HIC: high-income country; LMIC: low and middle-income country. #: the study website (www.bris.ac.uk/alspac/researchers/data-access/data-dictionary) contains details of all the data that are available through a fully searchable data dictionary.

Entebbe in Uganda (table 1), with a range of prevalence levels and exposures, and a likely range of phenotype distributions. Comparing findings from HICs and LMICs is not only of interest in itself, but is relevant to the assessment of whether associations seen in HICs (*e.g.* atopy and asthma) are causal or merely incidental. This is important both globally and within Europe, since Eastern and Southern Europe include middle-income countries which currently have very low asthma prevalence, but where prevalence is increasingly rapidly as these countries become more "westernised" [46].

Recruitment

Children are the most appropriate group for this study, since they are less likely to have undergone changes due to long-term asthma medication or occupational exposure. In four centres most participants will be children and adolescents aged 12–16 years, with an age range from 8–20 years. The ALSPAC [47] participants will be in their mid-20s and thus will be older than those in the other centres; however, this centre will add critical additional information because biological samples and information on asthma status are also available for earlier years and can be related to the subsequent phenotype distribution. Furthermore, the Wellington centre will provide another high prevalence centre in a HIC and will be focused on both children and adolescents. The specific findings in children in Wellington can thus be compared with those from children in LMICs and with those in young adults in ALSPAC.

In each centre, we will recruit a minimum of 200 asthma cases and 50 non-asthmatic subjects. The focus will be on defining categories within the groups of asthmatic subjects; however, a comparison will be made with the non-asthmatic subjects in each centre in order to check that the derived asthma phenotypes do in fact distinguish asthmatic cases (of different types) from non-asthmatic cases.

The recruitment methods will differ by centre. For example, the UK centre will recruit asthmatic subjects already identified through the ALSPAC study [25]. In other centres/countries, asthmatic subjects will be recruited through follow-up with participants from existing birth or other population-based cohorts and from schools. The comparison group of non-asthmatic subjects will be chosen from the same sources (*i.e.* other participants in the ALSPAC study or other children without asthma from the same school). The inclusion and exclusion criteria for the study are shown in Table 2.

Identification of asthmatic and non-asthmatic subjects

Asthmatic cases will be identified as those with symptoms of asthma and/or use of asthma medication in the past 12 months using the International Study of Asthma and Allergies in Childhood (ISAAC) questionnaire or, in adults, the European Community Respiratory Health Survey (ECRHS) questionnaire (the same key questions occur in both). Non-asthmatic cases will be identified as having no previous or current history of asthma using the ISAAC and ECRHS questionnaires.

Data collection

Information will be collected using standardised methods and operational procedures. These will include risk factor questionnaires and clinical characterisation, as well as blood, induced sputum and nasal lavage samples. We will repeat the sputum and nasal lavage samples after approximately 3 months in a sub-group of asthmatic subjects.

TABLE 2 Inclusion and exclusion criteria		
Criteria	**Cases**	**Controls**
Inclusion	Symptoms of asthma and/or use of asthma medication in the past 12 months	No previous or current history of asthma
Exclusion	Chronic disease apart from asthma[#] Pregnancy	Chronic disease[#] Pregnancy
Postpone appointment	Acute exacerbation of asthma Symptomatic respiratory infection in the last 4 weeks	Symptomatic respiratory infection in the last 4 weeks

[#]: people with the following chronic diseases should be excluded: chronic obstructive pulmonary disease, bronchiectasis, cystic fibrosis, any form of heart disease (including hypertension/high blood pressure, congenital heart disease or if taking heart medication), any other chronic lung or chest problems, tuberculosis, HIV.

Risk factor and symptom questionnaire

The risk factor and symptom questionnaire will collect information about current living conditions and other risk/protective factors (most of this information has already been collected prospectively for ALSPAC). Questions on the frequency and severity of clinical symptoms, and on the use of asthma medications, will also be included based on the ISAAC Phase II (asthma management) and Phase III (environmental risk factors) modules. Additional questions on current asthma control will be added based on the Asthma Control Questionnaire (ACQ; UK/English version, Sept 2010, Mapi Research Institute) by JUNIPER et al. [48]. In addition to the core questionnaire, there are three optional questionnaire modules on stress, nutrition and exposure to farm animals which the investigators at each centre can choose to include.

Skin prick tests

Skin prick tests (SPTs) will be carried out according to a well-defined protocol [49, 50] with histamine and saline being used as positive and negative controls, respectively. Fifteen minutes after testing, a mean wheal size of 3 mm or greater (once reaction to the negative control has been subtracted) will be considered positive. SPT positivity will be defined as a positive SPT to at least one of a panel of at least eight commercially available allergens, including house dust mite (*Dermatophagoides pteronyssinus*), tree pollen mix, grass pollen mix, cat and dog dander, *Alternaria tenuis*, Penicillium mix, plus locally relevant allergens (table S2).

Lung function testing

Lung function testing will be conducted according to American Thoracic Society (ATS) criteria using portable ultrasonic spirometers (the EasyOne® device from ndd Medical Technologies is the recommended tool, although others are also acceptable). All spirometers will be tested regularly (by the supplier) and calibration checks will be conducted weekly by research staff when devices are in use. Lung function testing will be carried out with participants in a sitting position, with three reproducible forced expiratory manoeuvres being performed to determine the following parameters: forced vital capacity (FVC), forced expiratory volume in 1 s (FEV_1), FEV_1/FVC, maximal mid-expiratory flow (MMEF) and peak expiratory flow (PEF).

Exhaled nitric oxide fraction

Exhaled nitric oxide fraction (F_{eNO}) is an optional procedure for this study. If centres choose to include it, they should use the NObreath monitor (Bedfont Scientific Ltd, Maidstone, UK).

Blood samples

Blood samples will be collected using standardised operational procedures. Atopy will be defined on the basis of allergen specific IgE (sIgE) given the dissociation between sIgE and SPT positivity in LMICs. We recognise the possibility of cross-reactivity of IgE to common carbohydrate determinants rather than to truly allergenic proteins, especially in centres in LMICs [51, 52]. A 15-mL sample of peripheral blood will be collected from each participant and added to two 5-mL ethylenediamine tetraacetic acid (EDTA) tubes (one for a full/complete blood count and one for plasma/DNA testing) and one 5-mL serum tube for IgE analysis. Blood samples will be processed within 2–4 h and stored in aliquots at −80 °C.

Sputum induction

Sputum induction will be conducted using a standardised protocol involving saline inhalation that we have used previously [14] and which is adapted from the protocol developed by GIBSON et al. [53]. We will follow the guidelines of the European Respiratory Society (ERS) Task Force working group on sputum induction where possible [54]. Participants will be pretreated with 200 mg salbutamol and then administered 4.5% hypertonic saline by oral inhalation using an ultrasonic nebuliser for increasing intervals (from 30 s to 4 min), up to a total of 15.5 min. At the end, participants will produce a sputum sample into a sterile plastic container and the induced sputum will be processed within 2 h of collection. At least 100 µL of sputum plugs should be selected from the sample and the volume measured. Dithiothreitol should then be added to the selected pellet and the mixture dispersed before being filtered through a 60-µm nylon mesh. The resulting suspension should be centrifuged and the supernatant removed and stored at −80 °C for subsequent laboratory analysis. The pellet should be resuspended and cell viability (trypan blue exclusion) and total cell count determined using a haemocytometer. Cytospins of the suspension should be prepared and stained (May-Grünwald Giemsa) to identify leukocytes. A differential cell count of 200 nonsquamous sputum cells will be conducted using light microscopy. Sputum supernatants will be analysed for markers of allergic inflammation or innate immunity (*e.g.* neutrophil elastase, eosinophil cationic protein (ECP), histamine, prostaglandin D2 (PGD2), IL-1β, IL-13, interferon-γ (IFN-γ) and IL-8) and markers associated with airway remodelling (*e.g.* matrix metalloproteinase-9 (MMP-9), tissue inhibitor of metalloproteinase 1 (TIMP-1), transforming growth factor-β (TGF-β) and

angiopoietin). We will also measure markers of neural involvement (*e.g.* acetylcholine, neurokinin A, Substance P and neurotrophins). Some markers of airway remodelling (*e.g.* MMP-1 and TIMP-1) and neural involvement (*e.g.* brain-derived neurotrophic factor (BDNF) and nerve growth factor (NGF)) will be measured in serum.

The key outcome from sputum will be the differential cell count. Other markers will be measured for small groups of participants initially to establish the proportion with detectable levels. The final list of measured markers will be available from the authors.

Nasal lavage

The focus will be on the findings from the sputum (lower airways); however, we will also conduct nasal lavage (upper airways) and compare the findings, since there is still some uncertainty as to their comparability and whether nasal lavage results (which are easier to collect) can be used as a surrogate for sputum results. Nasal lavage will be conducted using atomisers to instil 2 mL of saline (or PBS) into each nostril followed by blowing strongly through the nostril to collect the sample in a sterile container [55]. Cytospin slides will be prepared and differential cell counts conducted.

Sub-study of repeatability

In each centre, a sub-group of 50 asthma cases will be sampled again after 3 months in order to assess the repeatability of the sputum and nasal lavage cell counts.

Airways microbiome and epigenetics

Where sufficient sputum sample is obtained, aliquots will be taken for microbiome and epigenetic analysis and stored at −80 °C. The specific details of these analyses will be determined by the availability of tests at the time, as the methods are rapidly changing in these areas.

Gut microbiome

Stool sample collection and storage is optional (four of five centres are collecting or planning to collect these samples). This will enable a comparison of gut and airways microbiome in a future sub-study (subject to funding).

Other potential analyses

If there is sufficient material remaining after the core analyses described above, we will consider additional analyses of sputum and nasal samples, such as proteomics and mass spectrometry analyses (subject to funding).

Data analysis

Analyses will first be done separately for each centre. The findings will then be compared across centres and across groups of centres (*e.g.* high prevalence centres compared with low prevalence centres). A key concept here is "triangulation" (*i.e.* do we find comparable phenotype categories and does their distribution within the population of asthmatic subjects differ between centres?) [56]. Furthermore, do different statistical methods yield analogous phenotypes? If the derived phenotype categories are similar in the various centres, then a combined analysis will be conducted for the overall data set adjusting for centre.

We will first conduct descriptive analyses involving comparisons between asthmatic and non-asthmatic subjects. This will include information on demographic factors, lung function, asthma severity, age at onset, persistence, treatment, markers of airway inflammation and atopy. We will then use all of the information collected, clinical (including medication use), demographic, pathological and molecular, to define new asthma phenotypes by applying latent class analysis (LCA). This is a model-based clustering method for finding subtypes of related cases (latent classes) from multivariable categorical data. We will use Bayesian information criteria to decide on the number of phenotypes that provides the best fit and will explore the use of other state-of-the-art bioinformatics methods, particularly Bayesian machine learning and other "intelligent data" (*i.e.* "big data") methods. We will also evaluate techniques used widely in other omic-related fields to reduce the dimensionality into informative variables (*e.g.* self-organising maps (SOMs)). A characteristic feature of SOMs is their ability to map potentially complex non-linear interrelationships in multidimensional datasets into visually more approachable, typically two-dimensional planes of nodes. Individuals can be assigned to nodes, allowing the evaluation of between-nodal differences in demographic and behavioural factors, biomarkers and microbiome using summary statistics, univariate tests and regression models. This work therefore has considerable potential to advance beyond the state-of-the-art and to develop new key biomarkers for incorporation into definitions of phenotypes. The large datasets will be analysed using statistical packages (*e.g.* Mstat, R and Stata), publically available

mining algorithms (*e.g.* software from SHEN *et al.* [57]), as well as new methods implemented in high-level programming languages (*e.g.* C++ and Perl).

Once the latent classes have been defined we will repeat the initial descriptive analyses, comparing the clinical and immunological characteristics of the various asthma phenotypes and of the non-asthmatic subjects. The comparisons of the various asthma phenotypes will include symptoms, clinical severity and medication use. We will also compare the relative prevalences of the different phenotypes across centres.

The current study is not designed to investigate the effects of changing ICS dose on sputum eosinophil percentages. Indeed, such an approach would be extremely difficult to manage in a community setting across five different countries with varying degrees of healthcare coverage. However, as we are collecting data on asthma treatment at the time of assessment for all centres (specifically data on medication in the last 12 months in the core questionnaire and frequency of short-acting bronchodilator in the past week in the ACQ) and conducting repeat assessments in 50 participants from each centre, there is opportunity within the WASP study to: 1) assess stability of phenotypes over time with no changes in treatment in all centres; 2) assess stability of phenotypes over time with patient-initiated changes in treatment if/when they occur; and 3) assess stability of phenotypes in some LMICs (*i.e.* Uganda) after introduction of ICS (which is likely to occur due to clinical and ethical obligations). This will provide an indication of whether ICS use may have biased our analyses and to what extent. In addition, we expect that a sizeable proportion of all participants in HICs will not frequently use ICS. For example, in one recent study of 77 adolescent asthmatic subjects in Wellington, New Zealand (one of the participating centres), 63% reported using ICS in the last year, whilst only 39% reported using ICS in the last two weeks [14]. In this study there were no significant differences in sputum eosinophil percentages between ICS-treated and ICS-naïve participants. Therefore, given the size of the study (1000 asthmatic subjects across five centres) and a sizable proportion of non/infrequent ICS users in adolescents, we will have sufficient power to conduct analyses stratified by ICS use (frequent *versus* infrequent or no ICS use). This will provide further insight into whether ICS use may (partially) explain any differences in phenotype prevalence observed between centres. Finally, the absence of detectable eosinophilia in symptomatic, corticosteroid-naïve patients in several studies (including one large study of 995 ICS-naïve asthmatic subjects, in which 47% had no evidence of eosinophilia during repeated testing) suggests that NEA is a real phenomenon and not simply the result of ICS treatment [14, 58, 59]. Any bias due to ICS use may therefore be relatively small and, as noted above, we will assess this in stratified analyses.

The analyses of repeatability of sputum induction and nasal lavage will involve certain cell types and inflammatory markers in the sputum and nasal samples (eosinophils, ECP, neutrophils, myeloperoxidase (MPO), tumour necrosis factor-α (TNF-α), IL-8 and others). The initial analysis will involve comparing repeated measures based on these classifications. We will estimate the correlations of repeated measurements based on the continuous outcome of each individual marker of inflammation. The mean levels of cytokines and the more general inflammatory markers will be compared using standard t-test methods for comparing (geometric) means. Subsequently we will compare the sputum and nasal results by calculating correlations between outcomes using each method.

Study size and power

Based on previous research, we assume that each major phenotype will likely include an average of 50 cases in each centre. Therefore, in each centre, the study will have a greater than 80% power to detect a three-fold difference in eosinophil levels (or neutrophil levels) between the different phenotypes or between each phenotype and the non-asthmatic cases. It will also have a greater than 80% power to detect a two-fold difference in clinical characteristics between phenotypes (*e.g.* 50% compared with 25%). This is partly circular, since the phenotypes are based on these characteristics, but these estimates indicate that the study will have reasonable power to make comparisons between phenotypes and across centres. The study will also have a greater than 80% power to detect a relative prevalence ratio for a particular phenotype between two centres (*e.g.* 30% of cases *versus* 10% of cases). For an exposure with a population prevalence of 10%, the case-control analyses (adjusted for centre) will have a greater than 80% power to detect an odds ratio (OR) of 1.7; however, for an exposure with a population prevalence of 20% (*e.g.* pet ownership or parental smoking), they will have a greater than 80% power to detect an OR of 1.5. If a particular subtype of asthma which accounts for about 50% of cases (*e.g.* non-atopic asthma) is analysed separately, then, for an exposure with a population prevalence of 15%, the case-control analyses (adjusted for centre) will have a greater than 80% power to detect an OR of 2.0. However, for an exposure with a population prevalence of 20%, they will have a greater than 80% power to detect an OR of 1.8.

Ethics

Ethical approval for the study has been obtained from the London School of Hygiene and Tropical Medicine ethics committee (ref: 9776) and in all five study centres.

Discussion

Since the ISAAC surveys [60], the need for large international collaborative and integrative studies to understand the asthma epidemics in developed and developing countries has been clear. The WASP study aims to redefine asthma phenotypes in children, adolescents and young adults using a combination of clinical, demographic and pathological data. In particular, we expect that by combining data on inflammatory and other markers from sputum with detailed clinical information it will be possible to identify novel phenotypes. This study was developed and commenced prior to the recent Lancet Commissions Asthma Report [2]; however, our approach is highly consistent with the recommendations in this report and with previous recommendations [61].

The key and fundamentally unique feature of WASP is the inclusion of centres in LMICs as well as in HICs. This inclusion is likely to enable better characterisation of non-atopic phenotypes that may be more common in LMICs yet may also be present in a minority of asthma cases in HICs.

Given that the study centres are in a variety of settings across the world, it is essential to carefully standardise the study protocol so that we can be confident that the differences in phenotypes or distribution of phenotypes we observe are genuine and not influenced by protocol variations across study centres.

The study will initially be conducted in five centres, but we are publishing the protocol here to enable other centres around the world to carry out similar analyses using a compatible and standardised protocol. The new methods of disaggregating asthma phenotypes require large numbers of participants because the proportions of some phenotypes may be very low. As such, it will be important to collaborate to assimilate datasets for analysis by these new statistical methods [62].

References

1 Edwards MR, Saglani S, Schwarze J, *et al.* Addressing unmet needs in understanding asthma mechanisms: From the European Asthma Research and Innovation Partnership (EARIP) Work Package (WP)2 collaborators. *Eur Respir J* 2017; 49: 1602448.

2 Pavord ID, Beasley R, Agusti A, *et al.* After asthma: redefining airways diseases. *Lancet* 2018; 391: 350–400.

3 Douwes J, Gibson P, Pekkanen J, *et al.* Non-eosinophilic asthma: importance and possible mechanisms. *Thorax* 2002; 57: 643–648.

4 Pearce N, Pekkanen J, Beasley R. How much asthma is really attributable to atopy? *Thorax* 1999; 54: 268–272.

5 Pearce N, Douwes J, Beasley R. Is allergen exposure the major primary cause of asthma? *Thorax* 2000; 55: 424–431.

6 Douwes J, Boezen M, Pearce N. Chronic obstructive pulmonary disease and asthma. *In*: Detels R, Beaglehole R, Lansang MA, Gulliford M, eds. Oxford Textbook of Public Health, 5th Edn. Oxford, Oxford University Press, 2009, pp. 1021–1045.

7 Weinmayr G, Weiland SK, Bjorksten B, *et al.* Atopic sensitization and the international variation of asthma symptom prevalence in children. *Am J Respir Crit Care Med* 2007; 176: 565–574.

8 Barreto ML, Cunha SS, Fiaccone R, *et al.* Poverty, dirt, infections and non-atopic wheezing in children from a Brazilian urban center. *Respir Res* 2010; 11: 167.

9 Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* 2012; 18: 716–725.

10 Pekkanen J, Lampi J, Genuneit J, *et al.* Analyzing atopic and non-atopic asthma. *Eur J Epidemiol* 2012; 27: 281–286.

11   Douwes J, Pearce N. The end of the hygiene hypothesis? *Int J Epidemiol* 2008; 37: 570–572.

12   Moncayo AL, Vaca M, Oviedo G, *et al.* Risk factors for atopic and non-atopic asthma in a rural area of Ecuador. *Thorax* 2010; 65: 409–416.

13   Simpson JL, Grissell TV, Douwes J, *et al.* Innate immune activation in neutrophilic asthma and bronchiectasis. *Thorax* 2007; 62: 211–218.

14   Brooks CR, van Dalen CJ, Zacharasiewicz A, *et al.* Absence of airway inflammation in a large proportion of adolescents with asthma. *Respirology* 2016; 21: 460–466.

15   Hancox RJ, Cowan DC, Aldridge RE, *et al.* Asthma phenotypes: consistency of classification using induced sputum. *Respirology* 2012; 17: 461–466.

16   Fleming L, Tsartsali L, Wilson N, *et al.* Sputum inflammatory phenotypes are not stable in children with asthma. *Thorax* 2012; 67: 675–681.

17   Spycher BD, Silverman M, Kuehni CE. Phenotypes of childhood asthma: are they real? *Clin Exp Allergy* 2010; 40: 1130–1141.

18   Haldar P, Pavord ID, Shaw DE, *et al.* Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008; 178: 218–224.

19   Spycher BD, Minder CE, Kuehni CE. Multivariate modelling of responses to conditional items: new possibilities for latent class analysis. *Stat Med* 2009; 28: 1927–1939.

20   Deliu M, Yavuz TS, Sperrin M, *et al.* Features of asthma which provide meaningful insights for understanding the disease heterogeneity. *Clin Exp Allergy* 2018; 48: 39–47.

21   Anto JM, Bousquet J, Akdis M, *et al.* Mechanisms of the development of allergy (MeDALL): introducing novel concepts in allergy phenotypes. *J Allergy Clin Immunol* 2017; 139: 388–399.

22   Drews AC, Pizzichini MMM, Pizzichini E, *et al.* Neutrophilic airway inflammation is a main feature of induced sputum in nonatopic asthmatic children. *Allergy* 2009; 64: 1597–1601.

23   Ye WJ, Xu WG, Guo XJ, *et al.* Differences in airway remodeling and airway inflammation among moderate-severe asthma clinical phenotypes. *J Thorac Dis* 2017; 9: 2904–2914.

24   Martinez FD, Wright AL, Taussig LM, *et al.* Asthma and wheezing in the first 6 years of life. *N Engl J Med* 1995; 332: 133–138.

25   Henderson J, Granell R, Heron J, *et al.* Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax* 2008; 63: 974–980.

26   Moore WC, Meyers DA, Wenzel SE, *et al.* Identification of asthma phenotypes using cluster analysis in the severe asthma research program. *Am J Respir Crit Care Med* 2010; 181: 315–323.

27   Fitzpatrick AM, Teague WG, Meyers DA, *et al.* Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. *J Allergy Clin Immunol* 2011; 127: 382–389.

28   Siroux V, Basagana X, Boudier A, *et al.* Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J* 2011; 38: 310–317.

29   Savenije OE, Granell R, Caudri D, *et al.* Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *J Allergy Clin Immunol* 2011; 127: 1505–1512.

30   Weinmayr G, Keller F, Kleiner A, *et al.* Asthma phenotypes identified by latent class analysis in the ISAAC phase II Spain study. *Clin Exp Allergy* 2013; 43: 223–232.

31   Pavord ID. Asthma phenotypes. *Semin Respir Crit Care Med* 2012; 33: 645–652.

32   Simpson JL, Scott R, Boyle MJ, *et al.* Inflammatory subtypes in asthma: assessment and identification using induced sputum. *Respirology* 2006; 11: 54–61.

33   Haldar P, Pavord ID. Noneosinophilic asthma: a distinct clinical and pathologic phenotype. *J Allergy Clin Immunol* 2007; 119: 1043–1052.

34   Shaw DE, Sousa AR, Fowler SJ, *et al.* Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J* 2015; 46: 1308–1321.

35   Guiddir T, Saint-Pierre P, Purenne-Denis E, *et al.* Neutrophilic steroid-refractory recurrent wheeze and eosinophilic steroid-refractory asthma in children. *J Allergy Clin Immunol Pract* 2017; 5: 1351–1361.

36   Brooks CR, Van Dalen CJ, Harding E, *et al.* Effects of treatment changes on asthma phenotype prevalence and airway neutrophil function. *BMC Pulm Med* 2017; 17: 169.

37   Veres TZ, Rochlitzer S, Braun A. The role of neuro-immune cross-talk in the regulation of inflammation and remodelling in asthma. *Pharmacol Ther* 2009; 122: 203–214.

38   Pearce N, Beasley R, Burgess C, *et al.* Asthma Epidemiology: Principles and Methods. New York, Oxford University Press, 1998.

39   Chaudhuri R, McMahon AD, McSharry CP, *et al.* Serum and sputum neurotrophin levels in chronic persistent cough. *Clin Exp Allergy* 2005; 35: 949–953.

40   Mostafa GA, Reda SM, Abd El-Aziz MM, *et al.* Sputum neurokinin A in Egyptian asthmatic children and adolescents: relation to exacerbation severity. *Allergy* 2008; 63: 1244–1247.

41   Hilty M, Burke C, Pedro H, *et al.* Disordered microbial communities in asthmatic airways. *PLoS One* 2010; 5: e8578.

42   Cowan DC, Cowan JO, Palmay R, *et al.* Effects of steroid therapy on inflammatory cell subtypes in asthma. *Thorax* 2010; 65: 384–390.

43   Demarche SF, Schleich FN, Henket MA, *et al.* Effectiveness of inhaled corticosteroids in real life on clinical outcomes, sputum cells and systemic inflammation in asthmatics: a retrospective cohort study in a secondary care centre. *BMJ Open* 2017; 7: e018186.

44   Bacci E, Cianchetti S, Ruocco L, *et al.* Comparison between eosinophilic markers in induced sputum and blood in asthmatic patients. *Clin Exp Allergy* 1998; 28: 1237–1243.

45   Jatakanon A, Lim S, Barnes PJ. Changes in sputum eosinophils predict loss of asthma control. *Am J Respir Crit Care Med* 2000; 161: 64–72.

46   Sozanska B, Blaszczyk M, Pearce N, *et al.* Atopy and allergic respiratory disease in rural Poland before and after accession to the European Union. *J Allergy Clin Immunol* 2014; 133: 1347–1353.

47   Boyd A, Golding J, Macleod J, *et al.* Cohort profile: the 'Children of the 90s'–the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013; 42: 111–127.

48 Juniper EF, Gruffydd-Jones K, Ward S, *et al.* Asthma Control Questionnaire in children: validation, measurement properties, interpretation. *Eur Respir J* 2010; 36: 1410–1416.

49 Wickens K, Lane JM, Fitzharris P, *et al.* Farm residence and exposures and the risk of allergic diseases in New Zealand children. *Allergy* 2002; 57: 1171–1179.

50 Weiland SK, Bjorksten B, Brunekreef B, *et al.* Phase II of the international study of asthma and allergies in childhood (ISAAC II): rationale and methods. *Eur Respir J* 2004; 24: 406–412.

51 Amoah AS, Obeng BB, Larbi IA, *et al.* Peanut-specific IgE antibodies in asymptomatic Ghanaian children possibly caused by carbohydrate determinant cross-reactivity. *J Allergy Clin Immunol* 2013; 132: 639–647.

52 Hamid F, Versteeg SA, Wiria AE, *et al.* Molecular diagnostics and lack of clinical allergy in helminth-endemic areas in Indonesia. *J Allergy Clin Immunol* 2017; 140: 1196–1199.

53 Gibson PG, Wlodarczyk JW, Hensley MJ, *et al.* Epidemiological association of airway inflammation with asthma symptoms and airway hyperresponsiveness in childhood. *Am J Respir Crit Care Med* 1998; 158: 36–41.

54 Paggiaro PL, Chanez P, Holz O, *et al.* Sputum induction. *Eur Respir J* 2002; 20: Suppl. 37, 3s–8s.

55 Soto-Quiros M, Avila L, Platts-Mills TA, *et al.* High titers of IgE antibody to dust mite allergen and risk for wheezing among asthmatic children infected with rhinovirus. *J Allergy Clin Immunol* 2012; 129: 1499–1505.

56 Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016; 45: 1866–1886.

57 Shen RL, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* 2009; 25: 2906–2912.

58 McGrath KW, Icitovic N, Boushey HA, *et al.* A large subgroup of mild-to-moderate asthma is persistently noneosinophilic. *Am J Respir Crit Care Med* 2012; 185: 612–619.

59 Gibson PG, Simpson JL, Saltos N. Heterogeneity of airway inflammation in persistent asthma: evidence of neutrophilic inflammation and increased sputum interleukin-8. *Chest* 2001; 119: 1329–1336.

60 Pearce N, Aït-Khaled N, Beasley R, *et al.* Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax* 2007; 62: 758–766.

61 Muraro A, Lemanske RF Jr, Hellings PW, *et al.* Precision medicine in patients with allergic diseases: airway diseases and atopic dermatitis-PRACTALL document of the European Academy of Allergy and Clinical Immunology and the American Academy of Allergy, Asthma & Immunology. *J Allergy Clin Immunol* 2016; 137: 1347–1358.

62 Custovic A, Ainsworth J, Arshad H, *et al.* The Study Team for Early Life Asthma Research (STELAR) consortium 'Asthma e-lab': team science bringing data, methods and investigators together. *Thorax* 2015; 70: 799–801.